

Automatic epilepsy detection using fractal dimensions segmentation and GP–SVM classification

Jakub Jirka¹
Michal Prauzek¹
Ondrej Krejcar²
Kamil Kuca^{2,3}

¹Department of Cybernetics and Biomedical Engineering, Faculty of Electrical Engineering and Computer Science, VSB – Technical University of Ostrava, Ostrava Poruba, Czech Republic; ²Center for Basic and Applied Research, Faculty of Informatics and Management, University of Hradec Kralove, Hradec Kralove, Czech Republic; ³Biomedical Research Center, University Hospital Hradec Kralove, Hradec Kralove, Czech Republic

Objective: The most important part of signal processing for classification is feature extraction as a mapping from original input electroencephalographic (EEG) data space to new features space with the biggest class separability value. Features are not only the most important, but also the most difficult task from the classification process as they define input data and classification quality. An ideal set of features would make the classification problem trivial. This article presents novel methods of feature extraction processing and automatic epilepsy seizure classification combining machine learning methods with genetic evolution algorithms.

Methods: Classification is performed on EEG data that represent electric brain activity. At first, the signal is preprocessed with digital filtration and adaptive segmentation using fractal dimensions as the only segmentation measure. In the next step, a novel method using genetic programming (GP) combined with support vector machine (SVM) confusion matrix as fitness function weight is used to extract feature vectors compressed into lower dimension space and classify the final result into ictal or interictal epochs.

Results: The final application of GP–SVM method improves the discriminatory performance of a classifier by reducing feature dimensionality at the same time. Members of the GP tree structure represent the features themselves and their number is automatically decided by the compression function introduced in this paper. This novel method improves the overall performance of the SVM classification by dramatically reducing the size of input feature vector.

Conclusion: According to results, the accuracy of this algorithm is very high and comparable, or even superior to other automatic detection algorithms. In combination with the great efficiency, this algorithm can be used in real-time epilepsy detection applications. From the results of the algorithm's classification, we can observe high sensitivity, specificity results, except for the Generalized Tonic Clonic Seizure (GTCS). As the next step, the optimization of the compression stage and final SVM evaluation stage is in place. More data need to be obtained on GTCS to improve the overall classification score for GTCS.

Keywords: genetic programming, adaptive segmentation, SVM, fractal dimensions, EEG

Introduction

Epilepsy is a form of neurological disorder manifested by seizures, which are not linked to direct causes such as head injury or drug abuse. Epilepsy seizure is caused by a sudden electrical discharge in the brain, which may be a result of exogenous stimuli in case of reflex seizures or endogenous stimuli caused by chemical imbalance. About 1%–2% of the world population suffers from epilepsy.¹ Also, 25%–30% of this group cannot be treated surgically or by drugs.² These people suffer from so-called “refractory epilepsy”, and devices and methods for automatic epilepsy detection are

Correspondence: Kamil Kuca
Center for Basic and Applied Research,
Faculty of Informatics and Management,
University of Hradec Kralove,
Rokitanskeho 62, Hradec Kralove, 50003,
Czech Republic
Tel +420 60 328 9166
Email kamil.kuca@fnhk.cz

of great importance to them and for their quality of life.³ The most severe consequence of the epilepsy disease is sudden unexpected death in epilepsy.³

Electroencephalography (EEG) signal processing for final ictal, interictal activity is divided into the following steps:

1. Low pass signal filtration
2. Adaptive segmentation based on fractal dimension
3. Feature extraction and compression based on genetic programming (GP)–support vector machine (SVM) algorithm
4. Final stage SVM classification.

The most important part of signal processing for classification is feature extraction. Feature extraction is mapping from original input EEG data space to new features space with the biggest class separability value. Separation problem cannot be properly solved unless provided with proper features and their relationship consideration. Scientists also came to agreement that features are not only most important but also a most difficult task from the classification process. An ideal set of features would make the classification problem trivial.

In this paper, we introduce a novel method for EEG signal classification based on the methods of GP in combination with SVM confusion matrix, as the GP fitness function weighs for the tree structure individual quality, which represents the feature itself. Number of features is determined automatically by the novel function for the best feature selection. This part of an algorithm covers feature extraction and compression with the best feature selection in one stage. Reduced feature vector dimensionality also improves the performance of final SVM classification. This novel method is tested and proven to be more successful and effective for automatic epilepsy detection than most of the currently presented methods.

Approximately 1/100 of people suffer from a seizure attack once in their lifetime. The occurrence of an epileptic seizure is unpredictable and its cause is understood very little. The EEG signal which is used to present and test this novel algorithm is a signal obtained from the number of (usually 128, 34, 32, 19, 14) electrodes placed on a scalp. This signal contains the sum of electric activity produced by millions of neurons. This signal contains not only a lot of valuable information but also great amount of noise and useless information. Detection of ictal activity is very important in diagnosis and therapy.⁴

Visual analysis of the long-term EEG recording by the neurologist is very time consuming, and development of an automatic epilepsy detection algorithm had been analyzed in a number of studies.

Materials and methods

Automated epilepsy detection

Since the beginning of automated epilepsy detection, this field had developed in two main directions:

1. Interictal spike detection or epilepsy spike detection and
2. Epileptic seizure analysis.

This paper presents method from the group of automated spike detection based on methods of GP combined with SVM confusion matrix.

GP

GP method was developed for finding the approximate or exact problem solution by Koza.⁵ GP method is based on an individual evolution in the population tree, where every individual represents certain solution to the problem. GP method looks for the best solution using Darwin's evolution theory, where the original line of individuals (generation) is changed and selected by means of mutation, crossover, and natural selection.

Methods of GP were already used to improve k-nearest neighbors classification method performance, where a decision tree was declared for every feature, which was the function of the feature itself.⁶ Methods based on GP were also used in image processing to classify specific images.⁷ Another application found its place in the new feature vector extraction that improved the overall performance of the kNN classification method.⁸ GP-based methods were also used in optimal feature vector search that improved the classification algorithm efficiency.⁹

The method introduced in this article creates new synthetic set of features instead of searching for the best feature set as in the abovementioned articles.

Problem definition

The method introduced here is being developed as a part of the complex system for environmental reflex seizure stimuli localization using a series of sensor fusion and complex analysis methods. This complex system is a part of the dissertation thesis "Biometrical data processing and analysis – exogenous epilepsy seizure source localization and classification".^{10,11} Main motivation of this thesis and its part described in this paper is to improve the quality of life of people suffering from seizures and initiate development in areas of automatic exogenous stimuli detection and localization which had been overlooked by both scientists and doctors, and current research state in this field is far beyond its current possibilities.

There is a clear link between environmental stimulants and reflex seizures as described by researchers, but there is

also an expressed unsatisfactory state of current research in this area.^{12,13} The developed algorithm criteria were efficient in terms of low computational cost and high classification accuracy. All the above-described issues were important to improve the current research state in the area of automatic seizure detection in environmental stimuli localization application.

Current automatic seizure detection approaches are trading computational cost for high accuracy.¹⁴ Highly accurate methods come with high computational cost, and highly efficient methods used, for instance, in Brain Computer Interface come with results of lower accuracy.¹⁵ This study uses linear machine learning method, SVM, in combination with GP to introduce an efficient method with high classification accuracy in automatic epilepsy detection.

Details of specific signal preprocessing, basic feature set extraction, and implementation of this newly purposed GP-SVM algorithm are described in the “Automatic EEG signal segmentation using fractal dimension measure” and “Automatic epilepsy classification GP-SVM method” sections. Tests on sample of specifically selected 42 test subjects and final evaluation of this algorithm are described in the “Testing” and “Results and discussion” sections, respectively. The final conclusion and future work are described in the “Conclusion” and “Future work” sections, respectively.

Automatic EEG signal segmentation using fractal dimension measure

Before the EEG signal can be analyzed, it needs to be separated into time fragments called epochs. Because this algorithm is analyzing and searching for epileptic (ictal) activity, it is effective to use such measure for adaptive segmentation algorithm, which performs the best in separating ictal and interictal signal.

State of the art

The frequency and amplitude of EEG signal are highly variable. This is given by the nature of the biological signal source itself. The final EEG signal contains, besides useful information, lots of biological noise (electrocardiography [ECG], electromyography, electrooculography) and artificial environmental noise (electromagnetic disturbances). For further processing, it is necessary to separate the signal into time segments called epochs. There are two types of segmentations: constant with the constant time fragment length and adaptive with variable time fragment length changing according to the nature of the signal. Constant segmentation is not described any further as results of this old technique are not satisfactory.

Adaptive segmentation looks for transient states in the signal. Signal transition in the case of epileptic EEG data is transition from interictal to ictal state.

Adaptive segmentation algorithm was first presented in 1986 by Silin and Skrylev using two sliding Fast Fourier transform (FFT) windows. The algorithm was based on difference in the sliding window values. New segment is detected if the difference crosses a certain threshold. This basic method was not very effective, thanks to the non-stationary nature of the EEG signal. This algorithm was later on updated by Varri who used differences in signal amplitude summations in two sliding windows.¹⁵ This algorithm became the basic algorithm for all other adaptive segmentations. There is a whole series of adaptive segmentation algorithms, namely, linear prediction method, autocorrelation method, FFT, fuzzy c-means, and fractal dimensions. Linear prediction and autocorrelation methods do not support multichannel segmentation that is required. FFT and two-window method is not suitable for highly non-stationary epileptic signal, and fuzzy c-means method has a high computational cost. Adaptive segmentation based on fractal dimensions uses only this feature to separate the signal into segments and its calculation is fast and effective.

Fractal dimension is a statistical index of signal complexity given by signal detail change to size change.¹⁶ In EEG signal processing, the fractal dimension feature expresses signal's non-stationary value.

Signal's fractal dimension is calculated using Katz's algorithm. When the fractal dimension value thus EEG signal complexity crosses a certain threshold (0.4) calculated from series of heuristically performed experiments the new signal segment is detected.¹⁷

The fractal dimension is calculated using Equation 1, where $n = \frac{L}{d}$, $L \dots$ is the total length of the curve calculated as an Euclidian distance between samples and $d \dots$ is the curve diameter defined as the maximum distance of the first sample with any other signal sample

$$D = \frac{\log_{10}(L)}{\log_{10}(d)} = \frac{\log(n)}{\log(n) \log\left(\frac{d}{L}\right)} \quad (1)$$

The adaptive segmentation value is then calculated according to Varri using the following equation:¹⁵

$$G_j = |D_{j+1} - D_j| \quad (2)$$

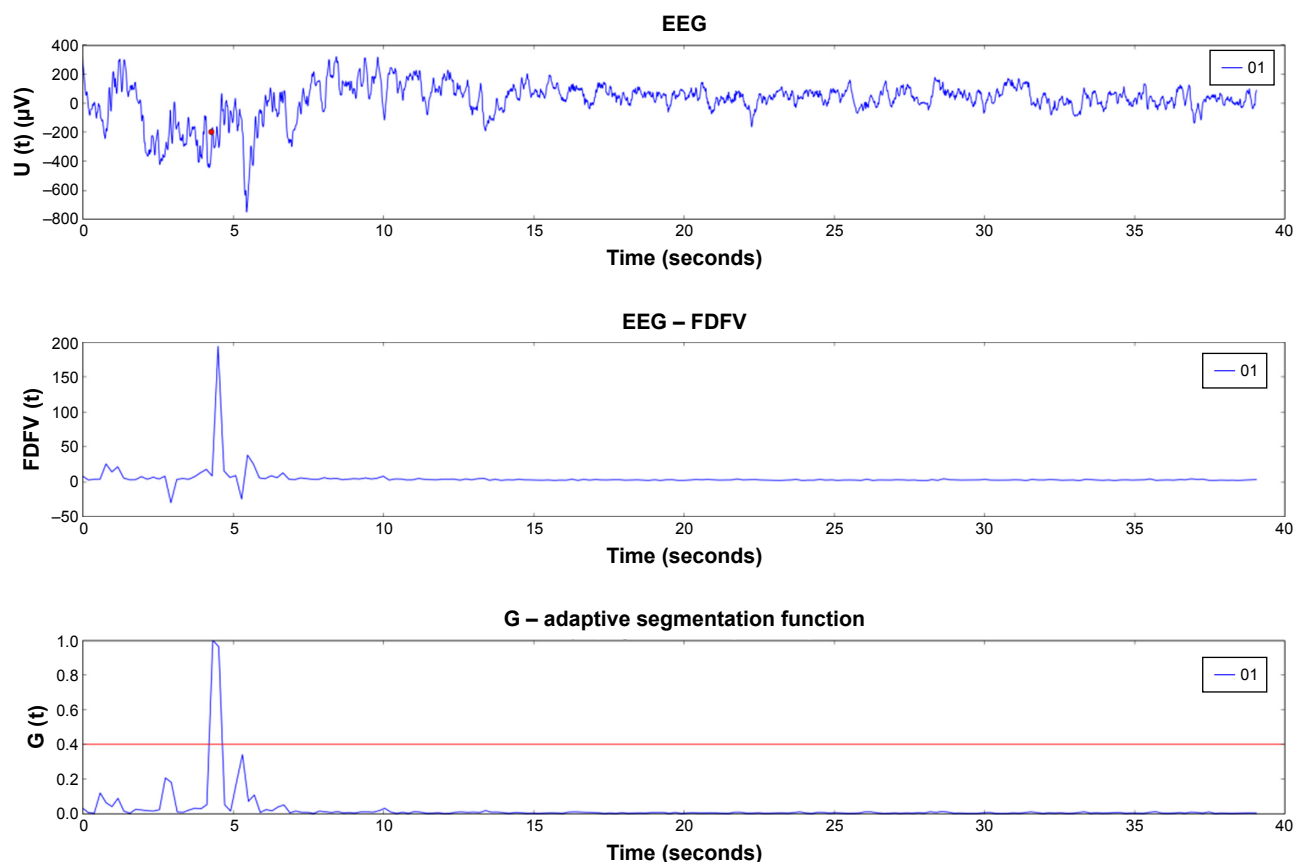


Figure 1 Result of automatic segmentation using fractal dimension on occipital area electrode signal from a patient with Jeavons syndrome.

Notes: A myoclonic epileptic seizure is detected and marked in the incipient segment. The upper graph represents the original EEG signal, the middle graph represents FDFV, and the lower one represents adaptive segmentation function. X-axis represents time in seconds. "U (t)" represents OI occipital electrode OI voltage data. "G" represents adaptive segmentation function as defined by Vari for the occipital electrode OI voltage data.

Abbreviations: EEG, electroencephalography; FDFV, fractal dimension feature vector.

The illustrative result of adaptive segmentation algorithm on data recorded from a patient with Jeavons syndrome is presented in Figure 1. The epilepsy segment is correctly detected just before the fifth second of the signal. The segment for further calculations has a 5-second overlap from the previous signal.

Automatic epilepsy classification GP-SVM method

Automatic feature extraction and compression

This paper presents automatic feature extraction and compression by creating new synthetic features of smaller dimensionality. Initial set of basic feature vectors is created by discrete wavelet transformation (DWT). The classification process using GP-SVM techniques (see Figure 2) can be summed up in the following steps:

1. Basic feature set extraction using DWT.
2. Synthetic feature set creation compressing original feature set and optimizing basic feature set.
3. Final classification using new feature set and SVM method.

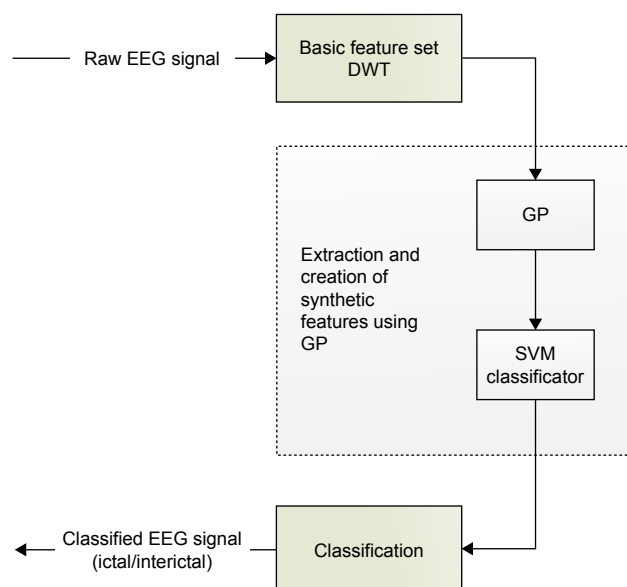


Figure 2 Schematic flow chart for set extraction and further processing (compression, classification using GP+SVM).

Abbreviations: DWT, discrete wavelet transformation; EEG, electroencephalography; GP, genetic programming; SVM, support vector machine.

Basic feature set

The first step described in the “Automatic feature extraction and compression” section is the basic feature set extraction using DWT. This transformation is used due to the fact that the EEG signal has a non-stationary character. Discrete form of wavelet transform uses quadrature mirrored set of finite impulse response filters where the low-pass filter’s output represents input signal approximation and high-pass filter’s output represents input signal detail components.

Every decomposed signal using DWT represents specific time frequency spectrum. A set of six features is extracted for every such decomposed information. These features are selected from area of time, statistical analysis, and information theory to extract the most important characteristics of the EEG signal. These features are as follows:

- Signal average value, which is represented by the signal’s amplitude average value (Equation 3).

$$\frac{1}{N} \sum_{i=1}^N x_i \quad (3)$$

- SD, which is the square root of the signal variance, where μ represents the mean value of a signal (Equation 4).

$$\sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu)^2} \quad (4)$$

- Signal energy represents the average value of the signal energy (Equation 5).

$$\frac{1}{N} \sum_{i=1}^N x_i^2 \quad (5)$$

- Signal curve length is also used in the fractal dimension-based segmentation and defines the length of the signal curve given by the sum of the horizontal line segments and is given by Equation 6.

$$\sum_{i=1}^N |x_{i+1} - x_i| \quad (6)$$

- Signal curve skewness is used as a quantification indicator for signal asymmetry distribution and is given by Equation 7, where μ and σ are the mean value and deviation of a signal, respectively.

$$\frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^3 - 3 \quad (7)$$

$$h[i] = \sum_{j=1}^M p[j] \log p[j] \quad (8)$$

$$p[l] = \frac{y[l]}{\sum_{k=1}^M y[k]} \quad (9)$$

Synthetic feature extraction and compression

The second stage of the presented algorithm is using combined methods of GP and machine learning SVM method.

Every individual in a genetic tree represents a new set of features, which is created by nonlinear transformation of the previous individual. These new features are put into SVM classifier that expresses the features of individual quality via its error function ξ , which is a part of the confusion matrix expressing the quality of classification using 10-fold cross-validation testing.¹⁸

In standard GP-based methods, one individual in a genetic tree expresses one result and in the end of the tree evolution, only one feature vector (result) is obtained. In this paper, we present a method that includes the F function in every genetic population tree which expresses the new feature vector. The F function has only one argument and one output which is a copy of the function input. The output value of function F is taken as one of the final synthetic feature vectors. Number of feature vectors and, thus, the compression level is defined by the number of F functions.

An example of such an F function is given in Figure 3. This population tree is fed with an input of ten values $x_1 \dots x_{10}$ and given two F functions (nodes) that form the new feature expressed as a combination of the population descendants. This algorithm uses the result of the F functions and not the result of the whole population tree evolution. The best descendant is given by the best basic arithmetic operation combination which is found by the maximum fitness value or minimum SVM error function value given by confusion matrix. Two new feature vectors will be created as shown in the picture: the first feature vector will be expressed as multiplication of input basic features $x_5 x_9$ and the second feature vector will be expressed as summation of $x_3 + x_8$.

SVM confusion matrix and GP fitness function

Most important part of the algorithm is the population member fitness evaluation. The value of this function is given by the misclassification value of the SVM classifier given by the confusion matrix obtained by the 10-fold cross-validation. The process of the fitness function evaluation

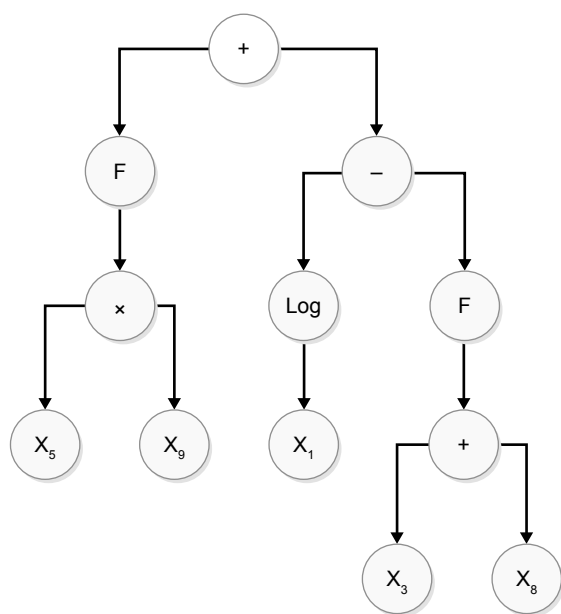


Figure 3 Population tree of suggested algorithm with two F functions (nodes) that make up new flagged vectors that are expressed by the offspring.

Note: In the illustrated example, two new flag vectors are created, the first being expressed as the sum of $x_5 \times x_9$ and the second as $x_3 + x_8$.

and best descendant search is shown in the block scheme in Figure 4.

Classification process can be summed up into the following steps:

1. Training input data are randomly separated into two groups: training (40%) and validation data (60%).

2. Evaluation of the population tree, best member selection, and new features extraction.
3. Trained SVM classifier classifies the validation data. Important value in this step is the value of the misclassified labels given by the confusion matrix. $\xi_i = N_v - N_s$, where N_v represents the total number of validation samples and N_s gives the number of correctly classified samples.
4. Steps 1–4 are repeated for every population member in search for the ones with the maximum fitness function value (minimum ξ_i value).

Data

The final conclusion and performance comparison table was obtained by testing the algorithm on 42 subjects in clinical and home environment setup.

EEG data were collected using devices from Alien Technik and Emotiv Epoc headset.

The work was approved by the Ethics Committee of Hospital Ostrava-Poruba. For our purposes, anonymized neurology data were obtained from the hospital, therefore, informed consent was waived by the ethics committee. Dataset was collected from 42 patients (27 male, 15 female) in 27 months (Feb 2012–June 2014). The patients were aged 4–62 years (mean: 23.83, SD: 17.43, median: 21.5). The study group included 21 children under age 18. Thirty-two patients were referred from the hospital neurological clinic, eight from Trigon association, and two were referred from

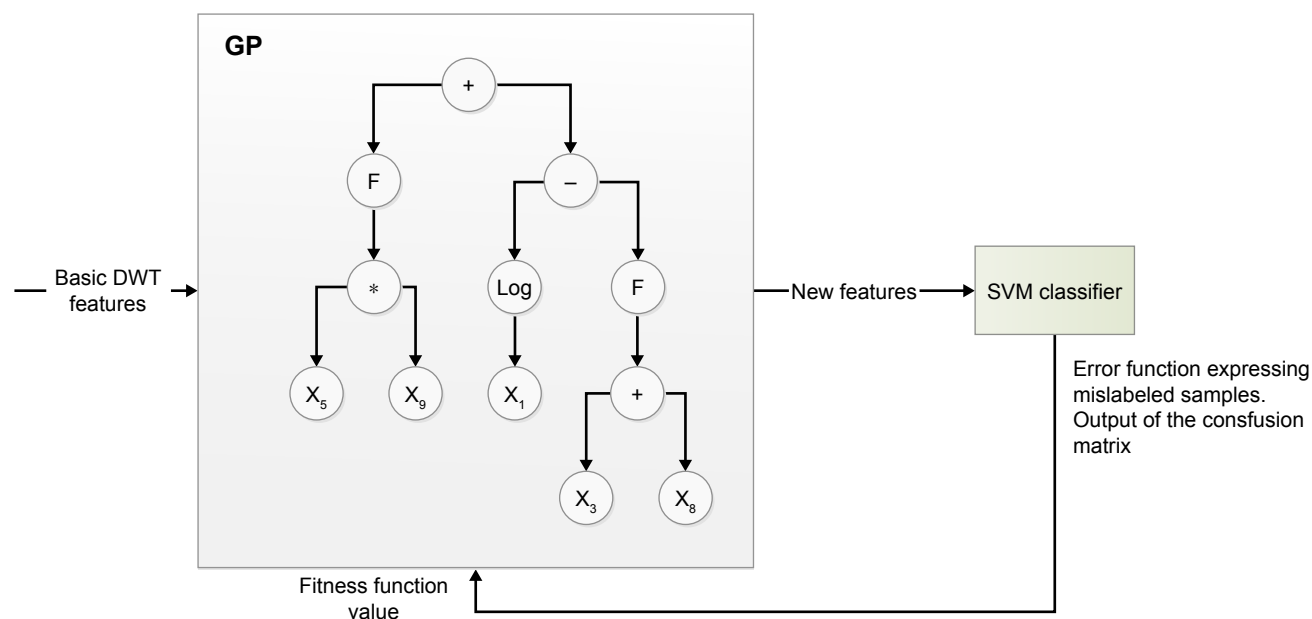


Figure 4 GP-SVM block scheme.

Abbreviations: DWT, discrete wavelet transformation; GP, genetic programming; SVM, support vector machine.

the GP. Thirty-six patients (85% of the group) were previously suffering from tonic-clonic episodes, of which 10 (24% of the group) were unwitnessed and not medically assessed and 26 patients (62% of the group) were medically diagnosed or EEG recordings were labeled to contain tonic-clonic seizure.

Clinical diagnosis

This section and Table 1 describe the clinical diagnosis of partial, generalized, and unclassified epilepsy types of patients. These data are based solely on the neurologist's evaluation provided either by the patient himself or by the neurologist upon patient's agreement. A total of 12 patients (29%) were classified with generalized epilepsy, 23 (54%) with partial epilepsy, and 7 patients (17%) were unclassified.

Testing

This section presents the results of the classification algorithm described in previous sections. Data were obtained in children from the neurological department of university Hospital Ostrava-Poruba and clients of Trigon association that provides support for people with social disabilities.

The patient whose results are shown in this section was diagnosed with photosensitive epilepsy, more specifically Jeavons syndrome.¹⁹ Data from this patient were fully classified and described by the neurologist. EEG signal record was fully labeled by the neurologist as well. These data contained

absence seizure, eyelid myoclonic seizures, tonic-clonic seizures, and myoclonic seizure.

Extraction of the basic feature vector

The basic features set is created by input signal decomposition using four-level DWT. Daubechies wavelet is used as the mother wavelet. Number of levels and type of mother wavelet were determined to be the best possible to discover most details and separate the EEG signal most appropriately according to its spectral composition.¹⁹ Wavelet filter bank frequency bands are specified for the sampling frequency of 128 Hz in Table 2.

Figure 5 shows the four-level wavelet decomposed signals for seizure-free signal. Figure 6 shows the EEG signal decomposition with myoclonic seizure.

Basic set of six features is calculated for every decomposed frequency sub-band according to features described in the "Basic feature set" section. Table 3 describes the input feature vector mapping with respect to decomposed sub-band and calculated feature.

Feature vector combination

Figure 7 shows results of SVM training phase defining hyperplane separating classified and non-classified samples using exponential radial basis function as defined in 10. Training samples of the population tree descendants combining the input features x_9+x_{10} are clearly distinguishable from the population member genetically combining the input features $x_{17}+x_{18}$, and thus, these two members are suitable for the new features.

These two members were found using fitness function value obtained from the confusion matrix, which is the output of the SVM classifier expressing classification quality.

An example of such confusion matrix can be observed in Figure 8. Elements of this matrix are used to express the quality of the classifier output. Diagonal elements represent the number of points correctly classified (predicted label = target true label) and off-diagonal elements express mislabeled

Table 1 Clinically diagnosed epilepsy types of patients

Seizure type	Number of patients
Generalized epilepsy	
Total	12
Myoclonus	2
GTCS with myoclonus	5
Photo-induced GTCS with myoclonus	3
Photo-induced GTCS with no other symptoms	1
Absence	1
Partial epilepsy	
Total	23
Simplex partial seizure	3
Complex partial seizure	5
GTCS with simplex partial seizure	14
Secondary generalized partial seizure	1
Unclassified	
Total	7
GTCS	3
Febrile seizures	2
Unclassified	2

Abbreviation: GTCS, generalized tonic clonic seizure.

Table 2 Frequency bands for four-level DWT decomposition

Signal label	Frequency band (Hz)	Decomposition level
w1	32–64	1
w2	16–32	2
w3	8–16	3
w4	4–8	4
a1	0–4	4

Abbreviation: DWT, discrete wavelet transformation.

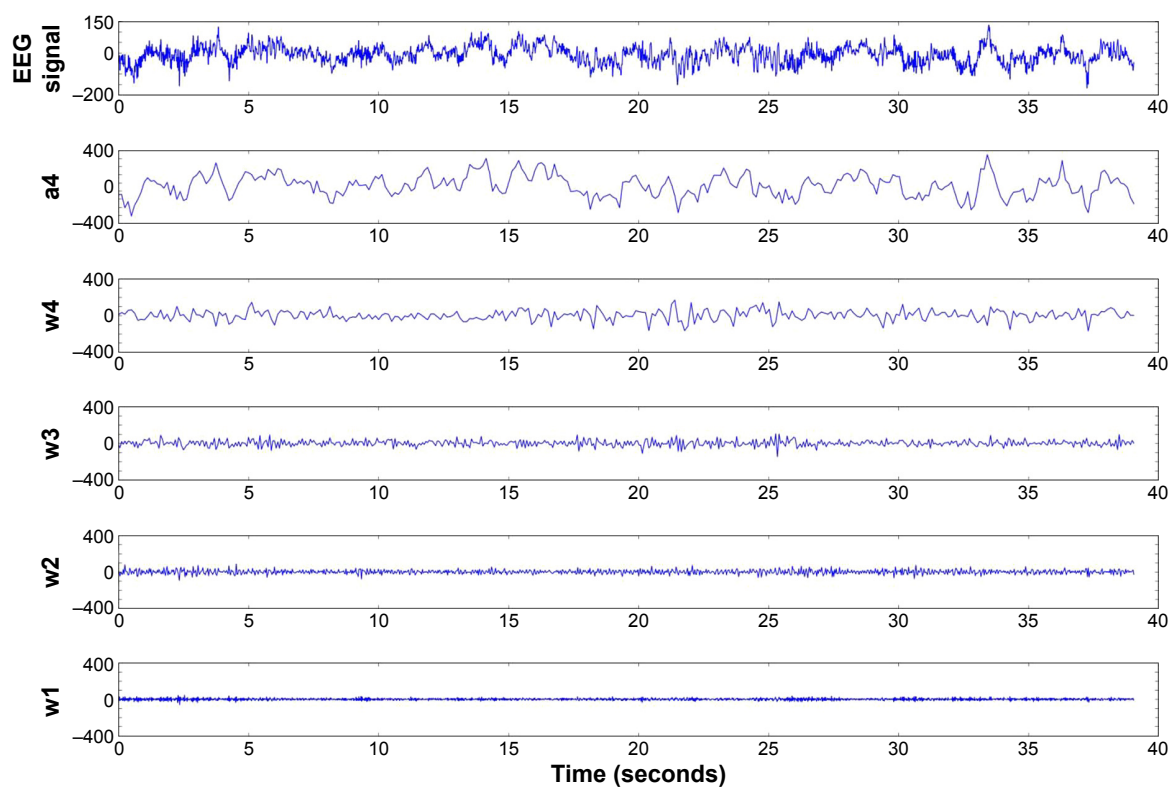


Figure 5 Four-level DWT decomposition for interictal EEG.

Abbreviations: DWT, discrete wavelet transformation; EEG, electroencephalography.

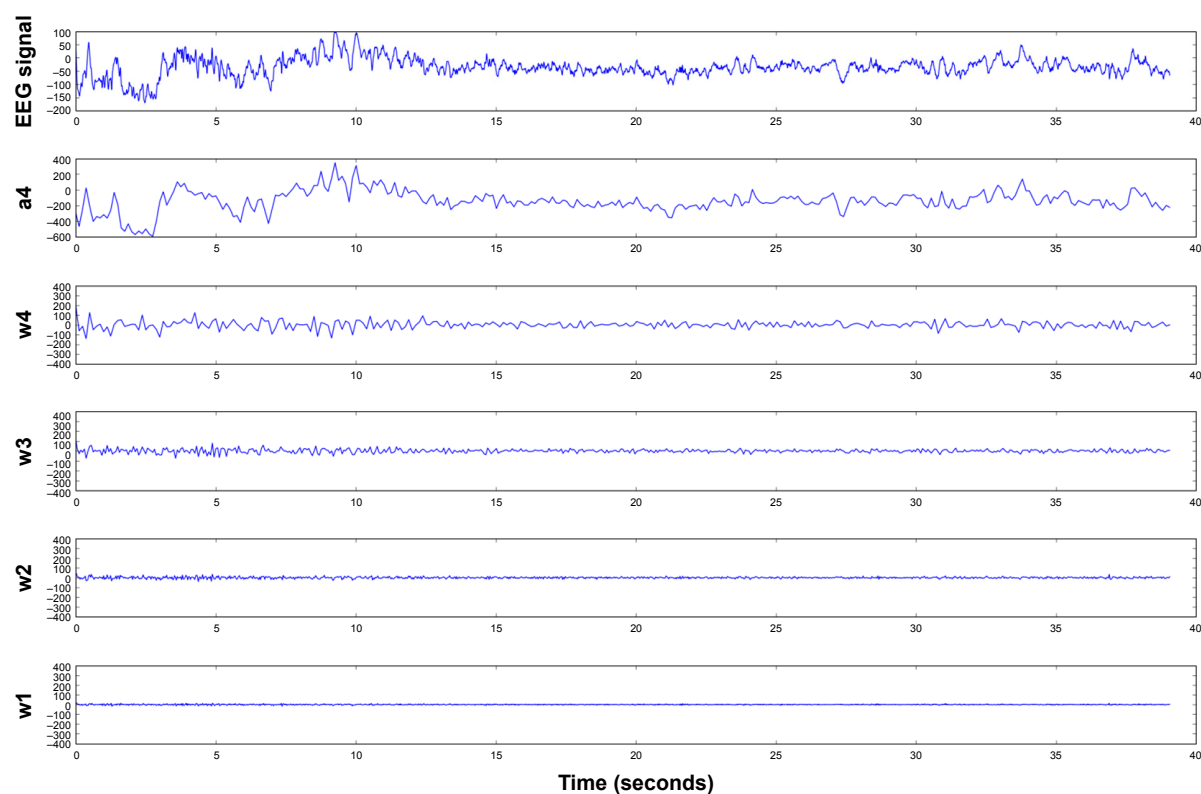


Figure 6 Four-level DWT decomposition for ictal EEG.

Abbreviations: DWT, discrete wavelet transformation; EEG, electroencephalography.

Table 3 Basic set feature vector mapping

	Frequency band (Hz)				
	32–64	16–32	8–16	4–8	0–4
Average value	x_1	x_7	x_{13}	x_{19}	x_{25}
SD	x_2	x_8	x_{14}	x_{20}	x_{26}
Energy	x_3	x_9	x_{15}	x_{21}	x_{27}
Curve length	x_4	x_{10}	x_{16}	x_{22}	x_{28}
Skewness	x_5	x_{11}	x_{17}	x_{23}	x_{29}
Entropy	x_6	x_{12}	x_{18}	x_{24}	x_{30}

samples. The higher the diagonal values mean, the better is the result of the classifier.

Results and discussion

Algorithm described in the “Automatic epilepsy classification GP–SVM method” section was tested on the dataset obtained from 42 subjects with 1–3 tests on each subject measured primarily over the night. Origin of the dataset is closely described in the “Data” section.

Algorithm results were evaluated using the following:

1. 10-fold cross validation technique X_{10} .
2. Statistical success rate evaluation of sensitivity X_E (classification success $<0; 1>$), specificity X_S (measure of the test ability to express the correct results – test quality), positive predictive value (PPV) – correctly detected events from all detected events, and negative predictive value (NPV) – falsely detected events from all detected events.

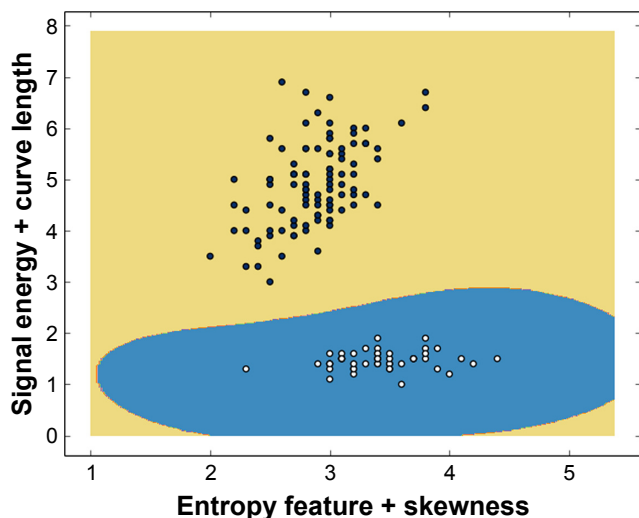


Figure 7 Training stage of population tree samples separation. Population tree descendants combining the input features x_9+x_{10} are clearly distinguishable from the population member genetically combining the input features $x_{17}+x_{18}$.

Notes: These two members (found using fitness function value) are suitable for the new features. Both the functions represented at X and Y axes are dimensionless.

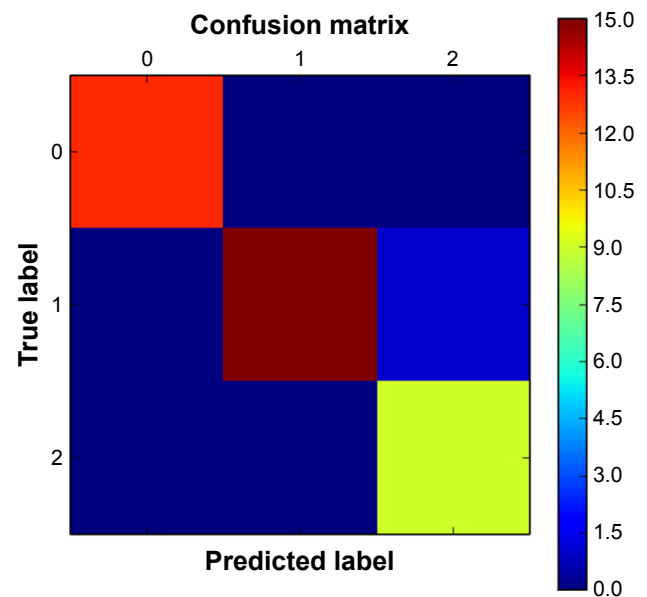


Figure 8 Confusion matrix for dataset shown in Figure 7.

Notes: Diagonal elements represent the number of points correctly classified (predicted label = target true label) and off-diagonal elements express mislabeled samples. The higher the diagonal values mean, the better is the result of the classifier.

10-fold cross validation technique can be summarized into the following steps:

1. Break data into ten sets of size $n/10$.
2. Train on nine datasets and test on one.
3. Repeat ten times and take the mean accuracy.

Table 4 summarizes the results of tests on a total of 77 EEG signal recordings using GP–SVM automatic epilepsy classification algorithm. Results are separated into groups defined by ictal and interictal class and seizure type labeled by the neurologist.

Basic feature set efficiency comparison with compressed feature set

The following test results display increased performance in epilepsy classification using basic feature set and compressed feature set with compression level 5 for a photosensitive patient. Table 5 displays the compressed basic vectors for occipital electrode O1. Table 6 presents results of the test comparing the performance results with PPV and NPV. A total of 356 events were recorded in the testing session.

Conclusion

This paper presents a novel method for automatic epilepsy detection using a combination of GP and SVM as a GP fitness function value and second stage stacking style

Table 4 Classification accuracy

Epileptic seizure type classification	$X_E \pm \sigma$	$X_S \pm \sigma$	PPV $\pm \sigma$	NPV $\pm \sigma$	$X_{10} \pm \sigma$
Ictal (myoclonic)	0.98 \pm 0.1	0.98 \pm 0.02	0.7 \pm 0.1	0.2 \pm 0.1	0.96 \pm 0.03
Interictal (myoclonic)	0.97 \pm 0.03	0.95 \pm 0.03	0.9 \pm 0.05	0.31 \pm 0.05	0.98 \pm 0.02
Ictal (GTCS)	0.95 \pm 0.5	0.83 \pm 0.02	0.7 \pm 0.2	0.23 \pm 0.2	0.84 \pm 0.1
Interictal (GTCS)	0.93 \pm 0.06	0.79 \pm 0.03	0.6 \pm 0.08	0.33 \pm 0.08	0.82 \pm 0.08
Ictal (tonic)	0.97 \pm 0.03	0.96 \pm 0.08	0.7 \pm 0.1	0.08 \pm 0.1	0.96 \pm 0.08
Interictal (tonic)	0.96 \pm 0.04	0.97 \pm 0.06	0.8 \pm 0.05	0.3 \pm 0.05	0.97 \pm 0.06
Ictal (atonic – mal absens)	0.99 \pm 0.03	0.98 \pm 0.01	0.9 \pm 0.03	0.2 \pm 0.03	0.97 \pm 0.02
Interictal (atonic – mal absens)	0.98 \pm 0.04	0.98 \pm 0.04	0.8 \pm 0.05	0.4 \pm 0.05	0.98 \pm 0.04

Abbreviations: GTCS, Generalized Tonic Clonic Seizure; NPV, negative predictive value; PPV, positive predictive value.

Table 5 Compressed feature vector with compression level 5

F_n	Compressed feature vector
F_1	$(x_1 + x_2)x_3$
F_2	$x_1 + x_9 + x_3$
F_3	$x_{15}x_{23} + \log x_{15}$
F_4	$x_{13} - x_5 + \log x_3$
F_5	$x_{15}x_6 - (x_8 + x_{15})$

classification for final ictal/interictal class. This novel method improves the overall performance of the SVM classification by dramatically reducing the size of input feature vector. According to the results, the accuracy of this algorithm is very high and comparable, and even superior to other automatic detection algorithms.¹⁴ In combination with the great efficiency, this algorithm can be used in real-time epilepsy detection applications. Test results of the algorithm can be seen in the “Results and discussion” section.

The results of the algorithm’s classification are of high sensitivity and specificity, except for the Generalized Tonic Clonic Seizure (GTCS). This bad performance might be the result of the small number of testing subjects with this specific seizure and also strong artifacts (moving, EMG) interfering with the original EEG signal. This had caused low PPV and X_{10} for this type of seizure.

Future work

This algorithm is currently being deployed and implemented on mobile device for real-time multimodal epilepsy detection using EEG emotive headset with a combination of acoustic, visual, and mechanic (accelerometer) modalities that support epilepsy detection. As the next step, the optimization of the compression stage and the final SVM evaluation stage is in place. More data need to be obtained on GTCS to improve the overall classification score for GTCS.

All the results and calculations described in this article had been performed using python and scikit libraries. Currently, this algorithm is being implemented on android phone device in Java to provide fully mobile support for the mobile non-clinical environment automatic epilepsy detection system.

Acknowledgments

This work was supported in part by the Center for Applied Cybernetics 3, Technology Agency of Czech Republic under project TE01020197, the Grant Agency of Excellence,

Table 6 Compressed and basic feature vectors-based classification performance comparison

	$X_E \pm \sigma$	$X_S \pm \sigma$	PPV $\pm \sigma$	NPV $\pm \sigma$
Basic features	0.873 \pm 0.051	0.876 \pm 0.113	0.437 \pm 0.235	0.2 \pm 0.1
Compressed features ($F_n=5$)	0.915 \pm 0.195	0.051 \pm 0.051	0.898 \pm 0.023	0.31 \pm 0.05

Abbreviations: NPV, negative predictive value; PPV, positive predictive value.

University of Hradec Kralove, Faculty of Informatics and Management, and the University Hospital Hradec Kralove – Long-term Development Plan. This work was also supported by the European Regional Development Fund in the Research Centre of Advanced Mechatronic Systems project, project number CZ.02.1.01/0.0/0.0/16_019/0000867.

Disclosure

The authors report no conflicts of interest in this work.

References

- Kandel ER, Schwartz JH, Jessell TM. *Principles of Neural Science*. New York: McGraw-Hill, Health Professions Division; 2000.
- Chapman D, Panelli R, Hanna J, Jes T. *Sudden Unexpected Death in Epilepsy: Continuing the Global Conversation*. Epilepsy Australia: Epilepsy Bereaved: SUDEP Aware; 2011.
- Stevenson A, Bardsley M, Billings J, et al; Whole System Demonstrator Evaluation Team. Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. *BMJ*. 2012;344:e3874.
- Subasi A. Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients. *Expert Syst Appl*. 2005;28(4):701–711.
- Koza JR. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. London and Cambridge, MA: MIT Press; 1992.
- Raymer ML, Punch WF, Goodman ED, Kuhn LA. *Genetic Programming for Improved Data Mining: Application to the Biochemistry of Protein Interactions, in Proceedings of the First Annual Conference on Genetic Programming*; Stanford, CA: MIT Press; 1996:375–380.
- Tackett WA. *Genetic Programming for Feature Discovery and Image Discrimination*; Urbana, IL: ICGA; 1993:303–311.
- Bot MC. Gen Program. Feature extraction for the k-nearest neighbor classifier with genetic programming. *Gen Program*. 2001;2038:256–267.
- Sabeti M, Katebi S, Boostani R. Entropy and complexity measures for EEG signal classification of schizophrenic and control participants. *Artif Intell Med*. 2009;47(3):263–274.
- Jirka J. *Biometrical Data Processing and Analysis – Exogenous Epilepsy Seizure Source Localization and Classification* [dissertation thesis]. Ostrava: VSB TU Ostrava; 2014.
- Ferlazzo E, Zifkin BG, Andermann E, Andermann F. Cortical triggers in generalized reflex seizures and epilepsies. *Brain*. 2005;128(Pt 4):700–710.
- Avanzini G. Musicogenic seizures. *Ann N Y Acad Sci*. 2003;999:95–102.
- Tzallas AT, Tsipouras MG, Tsilikakis DG, et al. Automated epileptic seizure detection methods: a review study. In: *Epilepsy – Histological, Electroencephalographic and Psychological Aspects*. Rijeka: In Tech; 2012:2027–2036.
- Lotte F, Congedo M, Lécuyer A, Lamarche F, Arnaldi B. A review of classification algorithms for EEG-based brain. *J Neural Eng*. 2007;4(2):R1–R13.
- Varri A. *Digital Processing of the EEG in Epilepsy* [licentiate thesis]. Tampere, Finland: Tampere University of Technology; 1988.
- Krajca V, Petráněk S, Patáková I, Váři A. Automatic identification of significant graphoelements in multichannel EEG recordings by adaptive segmentation and fuzzy clustering. *Int J Biomed Comput*. 1991;28(1–2):71–89.
- Hsu CW, Chang CC, Lin CJ. *A practical guide to support vector classification*; 2003. Available from: ntu.lib.ntu.edu.tw/bitstream/246246/20060927122852476378/1/freiburg.pdf. Accessed March 1, 2018.
- Harding GF, Jeavons PM. *Photosensitive Epilepsy*. London: MacKeith Press; 1994.
- Subasi A. Automatic detection of epileptic seizure using dynamic fuzzy neural networks. *Expert Syst Appl*. 2006;31(2):320–328.

Neuropsychiatric Disease and Treatment

Publish your work in this journal

Neuropsychiatric Disease and Treatment is an international, peer-reviewed journal of clinical therapeutics and pharmacology focusing on concise rapid reporting of clinical or pre-clinical studies on a range of neuropsychiatric and neurological disorders. This journal is indexed on PubMed Central, the 'PsycINFO' database and CAS,

Submit your manuscript here: <http://www.dovepress.com/neuropsychiatric-disease-and-treatment-journal>

Dovepress

and is the official journal of The International Neuropsychiatric Association (INA). The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.